March 11, 2005

To: Whom It May Concern

From: Kirsten Gallo and Steve Lanigan

Subject: Response to peer review of: Preliminary assessment of the condition of watersheds under the Northwest Forest Plan

We have reviewed the comments and suggestions submitted by 3 peer reviewers. We found their comments to be helpful in improving the report. The comments of the reviewers with our responses follow.

Reviewer 1
1. I am in general agreement with the method outlined in Reeves et al. 2004 in the use of distribution functions in assessing the diversity of watershed conditions It is a new and innovate concept with a lot of potential and hence its widespread application to assess watershed conditions across 25 million acres of national forests needs to be done carefully and have a robust foundation. The stated objective of the approach is to evaluate shifts in the distributions over time as a measure of improving or deteriorating conditions in watersheds. Hence, it is important to recognize (and incorporate into the analysis) the sensitivity of cumulative distribution functions (CDFs) to shifts, either in space or time. No such analysis is contained either in the Preliminary Assessment of the Condition of Watersheds Under he Northwest Forest Plan (PACW hereafter), in Reeves et al. (2004), or in any of the cited literature. First, CDFs are sensitive to the number of points used to create them and in particular the shapes of CDFs (and their range of variability) increases with the spatial (or temporal for that matter) sample size. Consequently, a CDF exhibiting a greater range of variability is less likely to be sensitive to land use induced variations in environmental conditions. For example, a CDF of forest ages obtained over a large area may contain burned areas (having young vegetation ages) contained within it and hence if the CDF was used as a "natural background condition" then some logging related reductions in stand ages would not stand out in comparison. While this is OK, the flip side is that if timber harvest was compared to a background CDF of stand ages constructed over a small area, then it would stand out as a major departure (this issue was mentioned in the Reeves et al. 2004 paper). Second, the ability to observe a shift in a CDF depends on the number of sample points used to create the distribution. For example, a handful of debris flows can cause shifts in a CDF of channel conditions when using a rather small sample size. However, a CDF containing a much large sample size shows very little shift over time. Therefore, a key step in the analysis proposed in Reeves et al. (2004) and applied in the PACW is to define the size of the sample sizes used to create the CDFs from which shifts are to be observed; this step is presently not included in the analysis.

    Response: The sensitivity analysis is an excellent suggestion. However, we don't feel that we have adequate data to conduct such an analysis. Ultimately the trend assessment will include in-channel data used in conjunction with the road and vegetation data. However, we have in-channel data for only 55 watersheds from a single time period. Without repeat surveys, we cannot assess our ability to detect trend. We did conduct an analysis of the number of sample sites as suggested, based on 50, 100, 150, 200, and 250 watersheds. Neither the CDFs nor our ability to detect changes in the CDFs changed with sample size. Thus, we decided to stick with the original 250 watersheds recommended by Reeves et al. (2004).

2. The use of large sample sizes in a CDF, such as stand ages either in upland or riparian areas, or total road miles across large land areas is going to make it difficult to observe statistically significant shifts in the CDFs. For instance, if road abandonment occurred within one HUC 5th watershed, it is likely that one would see a shift in total road miles in that

watershed. But due to economic constraints and the large sample size of existing roads, the relatively slow pace of road abandonment (due to economy) would make it unlikely to see shifts in the full distribution. This of course is what is reported in the PACW – no significant shifts occurred. This led to the inclusion of the caveat on pg. 17a stating that although no significant shifts occurred in the CDF, improvements nevertheless occurred due to road abandonment and riparian forest growth. If an analysis about how CDFs are sensitive to shifts had been done as part of the method development or in the PACW, then this result would have been understood during the study, and probably even before the study began.

> Response: See response to comment number 1.

3. The last point raises the issue about how the entire approach of shifts in CDFs is handled. The CDF approach lends itself very well to simulation modeling since it involves large spatial and temporal sample sizes over a landscape. For example, you could easily model or game scenarios about road abandonment by varying the sample size of roads (in a single watershed, populations of watersheds [landscapes], national forests, states, or regions) compared to actual or predicted pace of road abandonment. From this one could predict under what circumstances shifts in CDFs would occur and then plan accordingly – i.e., both helping to plan the pace of abandonment (although that will probably be dictated by economics) and helping define the appropriate sample size within the CDF. The same would hold true for riparian forest ages, although the problem is not as acute since forest growth is occurring everywhere at the same time. However, as determined in the PACW, significant shifts in the CDF of riparian forest ages were not observed. The inability to see shifts in CDFs over a short time frame (as mentioned in Reeves et al. [2004)] begs the question about why the analysis in the PACW was done in the first place. Here again, a simulation model may help. Simulating forest growth (and fires for that matter) on the CDF of forest ages is relatively trivial; there are many examples in the literature including in the U.S.F.S. CD on landscape dynamics (USFS 2002). One could have predicted a priori that no statistically significant shifts would have occurred in the CDFs of riparian forest ages simply by using a model. Moreover, such a simulation would also inform about when such shifts would be observable.

> Response: Running simulation models of road and vegetation change is not possible at this time, given the timeframe of this report. We did, however, determine the sensitivity of the models to road decommissioning and vegetation growth. We determined the time it would take for us to detect change given the rates of change observed in the last 10 years based on the sensitivity analysis conducted on the decision support models. Our next project is conducting an evaluation of the monitoring program. We added the simulation models as a method of arriving at the "desired" distribution as suggested by the review as a future program option.

4. In chapter 6 (pg. 33) under Emerging Issues, it states: " …the desired distribution of watershed conditions scores was not identified in the strategy. Consequently, we are unable to determine whether goals of the strategy were achieved." This limitation is also restated in a different form in the same section "…links between management and inchannel habitat and biological indicators – thus far, research has yet to determine how management activities in upslope and riparian areas affect fish and other aquatic and riparian dependent species." These are potentially fatal admissions since the watershed condition scores underpin much of the entire approach. It seems to me the construction of "evaluation models" (outlined in Appendix 3) is somewhat problematic since it is fighting an uphill battle to counter these stated limitations. Why not simply go with a predicted CDF of watershed attributes (under natural conditions) as the "evaluation model" [and also avoid in-channel measures of environmental conditions – see below]. For example, the CDF of riparian stand ages is given in Figure 18 of the PACW. That data could be contrasted with predicted CDFs of riparian forest ages that show the natural range of variability.

Response: The decision support models simply evaluate and aggregate all of the attributes in an index of watershed condition. We agree that the strength of the decision support models to estimate watershed condition would be greater if relationships between management and in-channel habitat were established and built into the model. However, the lack of these key pieces of information is not fatal. The lack of a "desired" distribution does not affect the utility of the decision support models or our ability to establish a distribution of watershed condition scores. The monitoring program's primary objective is to describe the condition of watersheds. Reeves (2004) discussed using simulation modeling or historical variation as "baseline" measures, but favored simply using the first period results, since the goal is to measure the effectiveness of the Plan. Although we recognize the utility of comparing the distribution of individual attributes to modeled distribution, we can not accomplish our objectives by only examining individual attributes as the reviewer described, nor do we have time to build the models the reviewer suggests within the established time frame.

5. Related to point #4 is the use of evaluation models (contained in the PACW) rather then a predicted CDF. For instance, Figure 22 of PACW shows a CDF of wood frequency. This is contrasted with the OR/WA Coast Province wood frequency evaluation model in Appendix 3 that shows 10 pieces/100 m is = -1 while 20 pieces/100 m is = +1. Wouldn't it be better to have an indication of the full range of variability from which to either establish "evaluation models". First, the model likely provides somewhat of a better estimate of the full range of variability in wood loading from which to compare present day data sets. Second, having a good simulation model of wood loading allows us to interpret the apparent low wood loading reflected in Figure 22 of PACW. For instance, does the low wood loading in Figure 22 match with the riparian stand ages in Figure 18?. I see there is no explanation or hypothesis regarding the low wood loading apparent in the data in the PACW. Use of a model might help. For instance, if low wood loading were encountered more frequently in eastern forests, it might be explained by the lack of recent fires, since the models predict that fires might be responsible for 50% of wood loading in drier forests (see Benda and Sias 2003). Although I don't disagree with the use of "professional judgment" in the formulation of evaluation models, its use (in the absence of constraining large data sets and/or simulation models) could lead to spurious models. For instance, the evaluation model for the Siskiyou/Klamath province indicates a value of 0 pieces/100m = -1 and 0.1 piece/100 m for +1. Based on wood inventories in approximately 100 km of streams in north central California (includes the Klamath system), the average wood loading (mostly on private land) is closer to 0.5 to 0.8 pieces/100m. Hence, the evaluation model for wood seems low.

Response: The evaluation criteria in the decision support models were developed based on the best available data. We recognize that these criteria should be treated as hypotheses and need to be validated. The simulation model the reviewer describes was developed for an individual watershed. We have no data to suggest that the model could be used with any degree of accuracy outside that watershed, and certainly not at the scale of the Northwest Forest Plan. The data needed to calibrate the model to extend its use to the Forest Plan area is not currently available. The reviewer uses large wood in the Klamath/Siskiyou province as an example of where simulation model results might improve the results of the decision support model. The reviewer suggests that the evaluation criteria for wood are low. However, 68 of 90 sites sampled by the monitoring program received an evaluation score of -1 (only 4 of the 90 sites received an evaluation score above 0), suggesting that wood levels in the areas we sampled are much lower than those the reviewer cites. In fact, we could argue that our criteria are too high. The disparity could be due to a difference in the minimum size of the wood counted or inclusion of log jams in the frequency calculation.

6. I will continue with the wood-loading example to bring up another point, that is, the use of random sampling across both watersheds and within watersheds (i.e., channel habitat surveys). Simulation modeling of LWD illustrates how spatial variability in LWD loading can

arise in a watershed. Hence, if one wanted to contrast actual wood storage with some measure of natural function, then where you measure it (or anything else for that matter) matters. This brings up the whole sampling protocol not to mention sample size. I am no statistician, but randomly sampling across watersheds (i.e., watershed selection) without an understanding about the spatial variability in watershed attributes from watershed to watershed or landscape to landscape that can impact the measured (or predicted) CDFs of those attributes seems somewhat questionable. In addition, the sample size of the streams (for in channel habitat attributes) seems slim at best, and again there is no apparent accounting for types of channels. For example, are similar types of channels sampled in all the watersheds? If so, then it would not be random sampling. Moreover, what about extrapolating results from the sampled watersheds to the others?

> Response: The generalized random stratified tessellation survey design was developed specifically for watershed monitoring, to allow us to use information from a subset of watersheds to make inferences about the population. As we conduct an evaluation of the monitoring program, we will use simulation models to test various sampling designs as the reviewer described. We will also conduct a test to determine the number of sites needed to describe the condition of watersheds.

7. As mentioned above, the amount of in-channel surveys seem to be very low considering the types of questions that are being posed (i.e., what is the overall aquatic habitat condition of a watershed and is it changing over time?). I think it is also highly questionable whether a "random sampling" approach should be used. The distribution of aquatic habitats, wood storage, pools, fines, bedrock etc. is not randomly distributed either in space or time. In addition, previous research has illustrated how intense spatial variability (not to mention temporal variability) precludes valid statistical analysis of change in channel morphology. My opinion is that it would be prudent to eliminate random, low intensity channel surveys from the entire approach for these reasons and others I can think of. It would be best to use hillslope and riparian surrogates to infer channel conditions throughout the plan area, and this could potentially be supplanted with intensively monitored watersheds in a few places. Surrogates could include ages of riparian forests, ages of upland forests, and frequency and magnitude of mass wasting etc.

> Response: As we conduct an evaluation of the monitoring program, we will be determining the number of sample sites needed to evaluate the condition of watersheds. We received criticism from managers because the assessment focused too heavily on vegetation and roads, and not enough emphasis was placed on the in-channel attributes. Therefore, we do not agree that we should drop the in-channel attributes completely and focus only on the upslope and riparian attributes, particularly since relationships between upslope and in-channel attributes have yet to be determined. That said, the sample design described, with a focus on upslope and riparian attributes, with intensive sampling in a few watersheds is among the alternatives that will be considered in the evaluation of the monitoring program.

8. The PACW's overall Conclusion: no statistically significant changes in the measured attributes occurred in the sampling period, although positive movement has taken place with respect to naturally growing riparian forests, the present absence of significant harvest (particularly in riparian areas), and road abandonment. This finding, that I agree with, would appear to be a defacto outcome of approach (use of distributions), their large sample size, and the small time step (a decade); see discussion above and some of the figures. Reeves et al. (2004) alluded to this issue. It is this fact that leads me to suggest using the CDF approach in a somewhat different way. That is, use simple province specific simulation models of vegetation (at least) to arrive at background CDFs (i.e., the evaluation model) and then future forest management could shoot for that as a target (or some reasonable facsimile of). This approach would have indicated the outcome (that was arrived at in the PACW) back in 1994 and it would also strongly recommend against detailed in channel measurements

simply because the models could not make accurate predictions (given our poor predictive ability regarding in channel and habitat attributes). Recall what is written in Chapter 4 of the PACW "...models had more trouble building the portion of the model that evaluates the in-channel data than the roads and vegetation portions. Stream channels have a very large range of natural variation, and determining what qualities any particular reach should have is difficult." The modeling approach would also be simpler, more transparent, and arguably more defensible and is consistent with Reeves et al. (2004). By the way, CLAMS (Oregon Coast Range) is leading the way in this direction.

> Response: We will consider the approach the reviewer has suggested among the program alternatives that we consider when we conduct the evaluation of the monitoring program. However, at this time, we do not agree that we should adopt this approach for this analysis. We currently don't have the resources to build the models that the reviewer suggests in the time frame established for this assessment. Nor do we feel that we should drop all of the in-channel information at this time (for reasons presented in #7).

9. In regard to the findings and conclusions of PACW, I think the report speaks to both the advantages and disadvantages of applying the general approach across landscapes to regions. Therefore, one aim of the PACW would be to evaluate the approach and make suggestions for future monitoring efforts.

> Response: We will be conducting such an evaluation in 2005.

10. Figure 5 caption is not Landsat.

> Response: The caption has been clarified.

11. Figure 56 and text on pg 25 having to do with stream layer resolution. A potential solution to this problem is to limit the proportion of the channel that you assess. Most of the error you describe is confined to headwater, 1st-order channels that can comprise the majority of stream length in a watershed. One could employ a "habitat threshold" based on a minimum flow (or perhaps gradient) that can support fish. This threshold can be used to parse down the network to the larger channels, say 3rd and higher order. That should get rid of the majority of the problem.

> Response: Excellent suggestion, we will consider this in future analyses.

12. Sampling stream attributes on pg. 17. The text states "...12% of the sample reaches had conditions scores equal to –1. These reaches either had high levels of fine sediment or were scoured down to bedrock." The question is were the "randomly selected" channels naturally prone to either high levels of fines or bedrock channel conditions? Without an answer, how is one to evaluate the result?

> Response: We don't know whether the channels were naturally prone to these conditions and know of no timely way to determine this. We are measuring watershed condition in an absolute sense, e.g., what is the habitat quality for fish, rather than relative to intrinsic potential (natural conditions). If these streams are naturally prone to fines/bedrock, the "natural" overall watershed score distribution will simply be lower. Since we are examining the distribution of watershed condition scores, and whether the scores increase through time due to management, and we don't expect all of the watershed to ever be in good condition, we don't think this lack of information is a critical flaw in the analysis.

13. Although it might be stated and I missed it, are the CDFs of attributes (upland forest ages, riparian forest ages, wood in streams, channel conditions etc.) province specific? My read of

this approach and based on some of the discussion above and the attached figures is that they should be.

> Response: All attribute evaluations were province specific. However, data from all provinces are presented in the figures. We specifically avoided presenting province-specific data because we didn't want to compare provinces. Because data from different provinces was evaluated using different models, the comparisons may not be fair.

Reviewer 2

14. My overall review of the document is that the presentation is too complex. The basic data that have been collected are very simple, and the concepts that need to be communicated are likewise fairly simple, but a degree of complexity is introduced in the analysis that inhibits communication of the facts. There are furthermore difficulties still apparent in the analysis that should have been cleared up before further peer-review took place. I would recommend a major revision of the document, including a revision of the analysis, figures, and outline before another review.

    A key part of the redrafting of this document is in looking at who the audience will be. Who will need to read this and understand it? As a scientist with a great deal of experience looking at cumulative density plots, non-linear responses, and multidimensional analyses, I found this document difficult to read and interpret. I expect that some of the key readers of this document will be congressional staffers with training in political science and policy analysis and little background in natural sciences or statistics. It is important that the key findings and interpretations be directly presented, while trends of lesser importance or "additional facts" should be set lower in the outline hierarchy. There are some aspects presented as part of results and discussion that might better be presented as independently written appendices to the report. For example assessments of the accuracy of the monitoring and sensitivity analysis. The presentation of the in-channel attributes is interesting but detracts from the overall message of the document because one cannot assess changes from data from a single point in time (more on that later). Focus on the comparison of road and vegetation data between the beginning of the plan and now!

    > Response: We redid many of the cumulative density plots as suggested by the reviewer, and moved the accuracy assessments and sensitivity analysis into appendices. We have received many comments, particularly from managers, that the analysis was too focused on roads and vegetation, and needed more information on the in-channel attributes. We feel that these attributes are important in setting the baseline distribution of watershed condition scores, which we are charged with completing. Reeves et al. (2004) recommended using the decision support model so we could integrate across attributes to describe watershed condition. Focusing only on the roads and vegetation attributes would not allow us to meet our other objectives in completing this assessment.

15. On page 17a a critical point about the analysis is revealed. This was a concern I was beginning to develop before I reached page 17. First, simple comparison of the distribution of condition at time 1 and time 2 while losing the information about which watershed had what value at a particular time is losing information over doing a test that takes into account the fact that repeated measures are being done, e.g. a pairing. There is no presentation of the magnitude of the changes outside of Table 5. Whoever authored page 17a apparently had a different set of table numbers. Looking at table 5, I see a very different story than is presented in the abstract or the rest of the paper. I would even guess that the author of page 17a sells short the available information content from table 5, although it is not immediately obvious what test would be appropriate for useful statistical information beyond what is presented. The non-symmetrical nature of the distributions would make a non-parametric test most appropriate.

Response: Page 17a contained results from the Wilcoxon test, which was based on the data in table 5. The test was rerun based on the actual data, rather than the binned data in table 5. Additional information on changes in individual watersheds has been added to the document.

16. My reading of table 5 is that "Most watersheds improved slightly, a few slightly more, and a few watersheds declined. All of the declines were related to loss of vegetation, as opposed to any road construction. Most of the increases in watershed condition related to vegetation growth. The condition of watersheds with respect to roads increased slightly in about 1/3 of the watersheds. A small number of watersheds saw more substantial improvement in road conditions, no watersheds saw a decline in road score." This seems like a much more common-sense, simple presentation, barely needing statistical support. The K-S test does not seem useful in this context.

Response: We restructured the results section to present the information more simply as recommended by the reviewer.

17. From this central statement of the overall findings, one can go to further analyses breaking down where the declines and improvement occurred – particularly with respect to land use designations. For the few watersheds with change magnitudes exceeding 0.3, a description of what occurred could be written, as it could for every watershed seeing a decrease in condition. Such breakdowns of land use designation or details of changes should only be described if there is useful feedback to the NW Forest Plan.

Response: The document has been modified as the reviewer suggested.

18. This relates to definitions of the goals of the monitoring and the definition of improvement given at the beginning of the document. My initial notes when reading page 5 (paragraph 2) were that we may not yet see more watersheds in good condition, and that if there is really a natural movement of watersheds among condition classes – up and down, then why should the number of watersheds in good condition be the key indicator. Why not just a higher distribution of scores? Why not just show the number of watersheds improving and declining and document the reasons – anthropogenic and natural. Using the number of "good" watersheds as an indicator of improvement seems to be inconsistent with the concept of dynamic watersheds. If we have every manager out there trying to push their watersheds into good condition, and battling natural processes to do it, are we really achieving our goals? In particular to use this sort of approach, you need to know the time scales of stream recovery from disturbance and the spatial scale of stream disturbances – relative to the 6th code huc sampling element you are using – in order to reasonably set targets. The scientific community does not know those.

Response: The approach the reviewer suggests – looking for higher watershed condition scores, rather than "good" watersheds – is actually what we intended. However, we clearly did not communicate our intentions. The section has been modified as the reviewer suggested to improve clarity.

19. Beyond that there is difficulty in the definition of "good" condition, which includes both measures of process and condition. There seems to be confusion between process and condition in the writing and conceptualization. Under natural conditions, processes operated to change conditions in watersheds, and conditions in individual watersheds varied from good to poor depending on disturbance and recovery processes and the basic capabilities of the habitat as constrained by topography, geology, climate, etc. In this monitoring paradigm, mostly reinforced by the condition scoring, one could conceivably have good process and poor condition (e.g. post fire) ranked as undesirable as poor process but good condition (e.g. under fire suppression). This may seem like a harmless mixing, but the insidious effect is to have form and process interchangeable to managers, when we really want the process to be

natural and the form to evolve according to the process. We should be monitoring to see how well we are setting up the landscape to behave naturally.

Response: We understand the difference between process and form that the reviewer describes. The questions the monitoring program is charged with answering are directed toward processes operating in watersheds, as is the Aquatic Conservation Strategy for the most part. We ran into trouble, however, when trying to assess the condition of watersheds, because it is much more difficult to measure and monitor processes than it is to measure and monitor form (current condition). Quite frankly, we're still trying to figure out how to monitor processes. We expect "good" processes to produce a certain number of watersheds currently in "poor" condition, but on the other hand there can be too much of a "good" thing. Our primary metric of the distribution of the population as a whole is designed to balance this situation.

20. Along these lines, I would suggest dropping most of the scoring idea until the discussion section (Chapter 4). The results section (Chapter 3) can just focus on the key parameters used in the watershed scoring, such as acreage in particular sizes of trees or road mileage in particular locations and present the changes in those characteristics or normalizations of them directly. This eliminates the value judgments placed on the scores on pages 18-20, such as a change of "only 0.1." Instead, actual changes in the watersheds can be presented objectively in the results section. In the discussion section, the weighting of the results as interpreted through regionally assembled scoring teams with non-linear rules can be introduced. This is the part where someone interprets the importance the changes stated objectively in the results section. In this interpretation, there is the opportunity to note, for example that, substantial changes in road density may yield no change in the score, because once you go beyond a certain density more roads don't really make it worse. At this stage there is also the opportunity to discuss the relative contributions of each component to the weighting and

Response: We disagree with the suggestion to drop the decision support model information in the results and use it only to provide context in the discussion. First, we don't think it is appropriate to introduce new information such as the models in the discussion section. Further, we think the decision support models provide a unique perspective on the factors that contribute to the condition of watersheds. We did restructure the results to focus first on the individual attributes then on the decision support model scores, and we removed statements that suggested a judgment was made as the reviewer suggested.

21. I have some question as to whether the measure of road density used is appropriate. For one, there is some problem with the estimate of stream density as discussed later, which greatly affects the stream mileage estimate. Second, consider a set area based road density…in a landscape with low stream densities, this network of roads would have less effect than it would in a landscape with a high stream density, yet the numerical relationship would be exactly opposite. If a densification algorithm is used that uses a constant support area for all locations, the index used is exactly the same as a per unit area index. What is the theoretical basis for using a per stream mile road density in place of a per unit area road density? Also more clarification of what is meant by densified and undensified stream networks is needed in the methods section.

Response: When we built the decision support models, we considered various measures of each of the attributes, including area based densities that the reviewer suggested. The riparian area calculation is based on the stream layer, therefore the problem that the reviewer describes when stream densities are low occurs regardless of the attribute measure used (riparian areas are small in areas with low stream density). We added information of densified and undensified streams in the methods as suggested by the reviewer.

22. The cumulative density plots are plain hard to read, even for someone experienced in reading them. In the sidebar describing how to read them, there is little advice on how to interpret them, just a further description of the axes. One of the key points is that the greatest density of values is represented by the steep parts of the curves. I would recommend dropping most of the figures showing cumulative density plots. If it is necessary to show distributions, the kernel density estimator plotting available in the R or S-Plus statistical packages is much easier for most people to read – lay person or scientist. It seems like the only need for the cumulative density plots was for visual validation of the K-S tests, which seem to under use the data.

   Response: We changed the cumulative frequency distribution figures as suggested.

23. Fig 12 has an error in the labeling of 2 Klamath/Siskiyou provinces

   Response: The error has been corrected.

24. Fig 11 should have labels like figures 12-14

   Response: The labels have been added.

25. Figures 54 and 55 have the axes flipped relative to each other and would be easier to use together if both had DOQ derived estimates on one axis and corporate derived on the other.

   Response: We changed the figures as suggested.

26. Some of the discussion in Chapters 5 and 6 seemed appropriate and useful. I have a few thoughts related to these. One thought is that these are discussion material based on the results and process of the monitoring done. They should be subheadings under the discussion as opposed to stand alone chapters.

   Response: The information in chapters 5 and 6 were provided in separate chapters for consistency with the other status and trend reports.

27. Why couldn't earlier channel survey data be used. There is a huge data base of aquatic habitat condition monitoring done in the 80s and 90s using Hankin-Reeves on many watersheds in the PNW … why couldn't that data have been used to evaluate condition improvement???

   Response: We did not use those data for three reasons: 1) data from the two time periods were not available for most of the 250 watersheds, 2) Region 6 changed their sampling protocols in the mid 1990s, therefore the data prior to 1995 is not comparable to the data collected after 1995, and 3) the quality of the data is not known.

28. Chapters 5 and 6 both suggest that a detailed road condition inventory might be a more useful measure of progress than remotely sensed locations. It comes at minimal cost and is the sort of data that can be used by local managers directly. It is a leading indicator as opposed to a trailing indicator, but is a more direct measure than road density and location. I think that this is a good recommendation in the hierarchy of efforts between detailed stream condition monitoring and remotely sensed road locations. Road condition and distribution has some quantifiable effects on habitat and is something that we can, more or less, directly control.

   Response: We are working with the agencies to collect this information and maintain it in a corporate database.

29. A number of times, they cite working with PNW scientists. I would suggest that they work with more than PNW scientists and work with scientists outside of Corvallis at times as well. There is a well developed group of hydrologic scientists there, but there are additional views of the world that could be of value to the process.

    Response: Excellent suggestion. Thank you.

Reviewer 3

30. Fundamentally, this chapter is not an assessment of watershed condition. It is a description of the development of measures of watershed condition and a monitoring plan for the Northwest Forest Plan. It would be unreasonable for anyone to expect more monitoring and assessment than the Plan has accomplished. This makes it critically important to identify a timeframe through which 1) a complete assessment of watershed conditions and 2) a reasonable measure of change in watershed conditions. <u>The omission of this timeframe is the greatest weakness of the draft chapter.</u>

    Response: This comment is related to comment #3 above. We incorporated a sensitivity analysis that describes the timeframe necessary to begin detecting significant changes as suggested.

31. This chapter needs to be very clear that this is an illustration of the analytical approach that will be used in the assessment of watershed conditions that will be conducted after completion of watershed monitoring under the Northwest Forest Plan. The Northwest Forest Plan has succeeded in developing an approach for monitoring watersheds for an extremely large and ambitious regional landscape management plan. This monitoring program in only two years has been able to monitor 55 watersheds out of 250 randomly selected watersheds. That is a major accomplishment in less than a decade. The true value of the Plan and the monitoring effort will be evident in the final analysis in the coming decade.

    Response: We added information to the summary and introduction to clarify the point made by the reviewer.

32. The Chapter needs to make it crystal clear from the start that there has not been sufficient time to collect and analyze all the relevant watershed data---especially the instream biological data. This is discovered on page 13. A simple paragraph following the first paragraph could point to the diorama and make it clear what this report contains and does not contain in the simplest terms.

    Response: A diorama and the paragraph suggested by the reviewer has been added.

33. The Chapter is clear that the monitoring is not complete and this is a preliminary assessment. The layout could be improved to make the preliminary nature and the finding more clear and distinct. Use of headings, matrices, visual indicators in matrices, tables and graphs, and other approaches for information illustration would greatly enhance people's understanding of the Chapter. A diorama would make the timeline of the creation of the Northwest Forest Plan, development of the monitoring approach, and the implementation of the monitoring effort.

    Response: We added headings and other information to improve clarity in the chapter as suggested.

34. Though watershed monitoring is incomplete and there has not been adequate time to document most changes, there are several important take-home messages that are almost hidden in the fourth paragraph of the Summary (page 3 and "33a"). These need to be highlighted more clearly (possibly as bolded bullets) at the start of the chapter. They are almost buried in the Summary. The Results, Discussion, and Summary can then explain

these findings more clearly. The relative conditions of upslope forests and riparian forests and relative conditions of federal forests and non-federal forests need to be emphasized because these will lay the foundations for future conclusions. A thorough discussion and rigorous critique of these findings now will make future assessments more effective.

Response: The findings have been highlighted in the summary and remaining chapters as suggested by the reviewer.

35. One of the apparent contradictions of the chapter is the finding that indices for pools and sediment were relatively high and indices for large wood were low. This discrepancy is not adequately explained and is certain to received critical attention. As the chapter notes, watersheds with known habitat problems exhibited relatively high indices for in-channel habitat. The monitoring team needs to carefully investigate this pattern. This also needs to be linked to biological responses, especially fish abundance and distribution.

Response: We currently do not know the reason for the unexpected relationships between pool and sediment indices and large wood. Our refinement of the monitoring process will explore the relationship(s) between attributes; consider how to address the lag time between management activities and impacts to channel condition; and determine what relationships exist between biological metrics and upslope, riparian, and in-channel conditions.

36. One possible problem is the "single index problem" and the lack of a river network framework that is a weakness of PacFish/InFish. The proportion of pools will change from headwater stream to large river in any river network. Using a single value for pool frequency as an index of condition results in a "middle numbers game". The index does not appropriate measure the expected pattern of pools along a river network and therefore generates numbers somewhere in the middle of the expected distribution. Thus the index is relatively insensitive to measuring condition or detecting change. Assuming network pattern can be incorporated, it also is important to measure major features, such as deep pools. A measure of all pools will not reveal important changes in the loss of deep pool or large pool habitats (see work of Sedell, McIntosh, and Reeves on historical trends and ranges of conditions).

Response: We agree that the use of the decision support models results focusing on averages, rather than a more realistic range of conditions. We view the decision support model refinement as an iterative process. As we gain more knowledge and data, we will incorporate more attributes such as deep pools, and context variables into the models to improve our ability to describe the condition of watersheds. This year, we will be conducting an evaluation of the monitoring program, during which we will be examining the within and between watershed variability in the in-channel attributes, and evaluating our ability to capture the necessary information about these attributes.

37. The in-channel habitat indices will be the link between riparian conditions and biological responses in aquatic ecosystems. I think the chapter should highlight anticipated measures to insure a rigorous link between riparian, in-channel habitat, and biological indices. The options listed under "Reconciling sample design with funding limitations" needs to take this requirement into consideration. The text was not clear on how the decisions about the four options would be made. One of the most difficult choices in the monitoring program has been the choice between a few focused measures versus and large number of measures that we hope can be measured. The chapter helps the reader understand the array of indices and the sampling program but it does not help the reader understand how the range of measurements will be focused and strengthened over the next 5-10 years.

Response: Additional information on the analyses that will be conducted as part of the refinement, and the criteria for selecting an alternative was added to the chapter.

38. Starting on page 49, a large volume of unnecessary material is inserted into the chapter. Numerous photographs of people and procedures are provided. I see no reason for most of these. The sidebars (except for the cdf explanation) also are not necessary. The length of the chapter could be reduced by at least 30% if not more.

Response: Although removing this information would certainly shorten the chapter, we feel that side bars and photographs provide insight on the kinds of restoration work that is being performed under the Northwest Forest Plan, and the work that the monitoring team is conducting. We chose not to remove this information from the document.